The Annals of Statistics
2008, Vol. 36, No. 5, 2153–2182
DOI: 10.1214/07-AOS539
© Institute of Mathematical Statistics, 2008

# A CLASS OF RÉNYI INFORMATION ESTIMATORS FOR MULTIDIMENSIONAL DENSITIES

By Nikolai Leonenko,[1] Luc Pronzato [2] and Vippal Savani

*Cardiff University, CNRS/Université de Nice–Sophia Antipolis and Cardiff University*

A class of estimators of the Rényi and Tsallis entropies of an unknown distribution $f$ in $\mathbb{R}^m$ is presented. These estimators are based on the $k$th nearest-neighbor distances computed from a sample of $N$ i.i.d. vectors with distribution $f$. We show that entropies of any order $q$, including Shannon's entropy, can be estimated consistently with minimal assumptions on $f$. Moreover, we show that it is straightforward to extend the nearest-neighbor method to estimate the statistical distance between two distributions using one i.i.d. sample from each.

## 1. Introduction.

We consider the problem of estimating the Rényi [33] entropy

$$H_q^* = \frac{1}{1-q}\log\int_{\mathbb{R}^m} f^q(x)\,dx, \qquad q \neq 1, \tag{1.1}$$

or the Havrda and Charvát [15] entropy (also called Tsallis [37] entropy)

$$H_q = \frac{1}{q-1}\left(1 - \int_{\mathbb{R}^m} f^q(x)\,dx\right), \qquad q \neq 1, \tag{1.2}$$

of a random vector $X \in \mathbb{R}^m$ with probability measure $\mu$ which has density $f$ with respect to the Lebesgue measure, from $N$ independent and identically distributed (i.i.d.) samples $X_1, \ldots, X_N$, $N \geq 2$. Note that $H_q^*$ can be expressed as a function of $H_q$. Indeed, $H_q^* = \log[1-(q-1)H_q]/(1-q)$, and

Received July 2007; revised July 2007.

[1] Supported by EPSRC Grant RCMT119.

[2] Supported in part by the IST Programme of the European Community under the PASCAL Network of Excellence, IST-2002-506778.

*AMS 2000 subject classifications.* 94A15, 62G20.

*Key words and phrases.* Entropy estimation, estimation of statistical distance, estimation of divergence, nearest-neighbor distances, Rényi entropy, Havrda–Charvát entropy, Tsallis entropy.

for any $q$, $d(H_q^*)/d(H_q) > 0$ and $[d^2(H_q^*)/d(H_q)^2]/(q-1) > 0$. For $q < 1$ and $q > 1$, $H_q^*$ is thus a strictly increasing concave and convex function of $H_q$ respectively and the maximization of $H_q^*$ and $H_q$ are equivalent. Hence, in what follows we shall speak indifferently of $q$-entropy maximizing distributions. When $q$ tends to 1, both $H_q$ and $H_q^*$ tend to the (Boltzmann–Gibbs-) Shannon entropy

$$H_1 = -\int_{\mathbb{R}^m} f(x) \log f(x)\, dx. \tag{1.3}$$

We consider a new class of estimators of $H_q$ and $H_q^*$ based on the approach proposed by Kozachenko and Leonenko [21] who consider the estimation of $H_1$; see also [11]. Within the classification made in [3], which also contains an outstanding overview of nonparametric Shannon entropy estimation, the method falls in the category of nearest-neighbor distances. See also [13]. When $m = 1$, the nearest-neighbor method is related to sample-spacing methods; see, for example, [41] for an early reference concerning Shannon entropy. It also has some connections with the more recent random-graph approach of Redmond and Yukich [32], who, on the supposition that the distribution is supported on $[0,1]^m$ together with some smoothness assumptions on $f$, construct a strongly consistent estimator of $H_q^*$ for $0 < q < 1$ (up to an unknown bias term independent of $f$ and related to the graph properties). For $q \neq 1$, our construction relies on the estimation of the integral

$$I_q = \mathbb{E}\{f^{q-1}(X)\} = \int_{\mathbb{R}^m} f^q(x)\, dx \tag{1.4}$$

through the computation of conditional moments of nearest-neighbor distances. It thus possesses some similarities with that of Evans, Jones and Schmidt [8], who establish the weak consistency of an estimator of $I_q$ for $m \geq 2$ and $q < 1$ under the conditions that $f$ is continuous and strictly positive on a compact convex subset $\mathcal{C}$ of $\mathbb{R}^m$, with bounded partial derivatives on $\mathcal{C}$. In comparison to Redmond and Yukich [32] and Evans, Jones and Schmidt [8], our results cover a larger range of values for $q$ and do not rely on assumptions of regularity or bounded support for $f$. For the sake of completeness, we also consider the case $q = 1$, that is, the estimation of Shannon entropy, with results obtained as corollaries of those for $q \neq 1$ (at the expense of requiring slightly stronger conditions than Kozachenko and Leonenko [21]).

The entropy (1.2) is of interest in the study of nonlinear Fokker–Planck equations, with $q < 1$ for the case of subdiffusion and $q > 1$ for superdiffusion; see [38]. Values of $q \in [1,3]$ are used by Alemany and Zanette [1] to study the behavior of fractal random walks. Applications for quantizer design, characterization of time-frequency distributions, image registration and indexing, texture classification and image matching etc., are indicated by Hero et al.

[16], Hero and Michel [17] and Neemuchwala, Hero and Carson [29]. Entropy minimization is used by Pronzato, Thierry and Wolsztynski [31], Wolsztynski, Thierry and Pronzato [45] for parameter estimation in semi-parametric models. Entropy estimation is also a basic tool for independent component analysis in signal processing; see, for example, [22, 23].

The entropy $H_q$ is a concave function of the density for $q > 0$ (and convex for $q < 0$). Hence, $q$-entropy maximizing distributions, under some specific constraints, are uniquely defined for $q > 0$. For instance, the $q$-entropy maximizing distribution is uniform under the constraint that the distribution is finitely supported. More interestingly, for any dimension $m \geq 1$, the $q$-entropy maximizing distribution with a given covariance matrix is of the multidimensional Student-$t$ type if $m/(m+2) < q < 1$; see [43]. This generalizes the well-known property that Shannon entropy $H_1$ is maximized for the normal distribution. Such entropy-maximization properties can be used to derive nonparametric statistical tests by following the same approach as Vasicek [41] who tests normality with $H_1$; see also [11].

The layout of the paper is as follows. Section 2 develops some of the motivations and applications just mentioned (see also Section 3.3 for signal and image processing applications). The main results of the paper are presented in Section 3. The paper is focused on entropy estimation, but in Section 3.3 we show how a slight modification of the method also allows us to estimate statistical distances and divergences between two distributions. Section 4 gives some examples and Section 5 indicates some related results and possible developments. The proofs of the results of Section 3 are collected in Section 6.

## 2. Properties, motivation and applications.

2.1. *Nonlinear Fokker–Planck equation and entropy.* Consider a family of time-dependent p.d.f.'s $f_t$. The p.d.f. that maximizes Rényi entropy (1.1) [and Tsallis entropy (1.2)] subject to the constraints $\int_{\mathbb{R}} f_t(x)\,dx = 1$, $\int_{\mathbb{R}}[x - \bar{x}(t)] \times f_t^q(x)\,dx = 0$, $\int_{\mathbb{R}}[x - \bar{x}(t)]^2 f_t^q(x)\,dx = \sigma_q^2(t)$, for fixed $q > 1$, is the solution of a nonlinear Fokker-Planck (or Kolmogorov) equation; see [38].

Let $X$ and $Y$ be two independent random vectors respectively in $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$. Define $Z = (X, Y)$ and let $f(x, y)$ denote the joint density for $Z$. Let $f_1(x)$ and $f_2(y)$ be the marginal densities for $X$ and $Y$ respectively, so that $f(x, y) = f_1(x) f_2(y)$. It is well known that the Shannon and Rényi entropies (1.3) and (1.1) satisfy the additive property $H_q^*(f) = H_q^*(f_1) + H_q^*(f_2)$, $q \in \mathbb{R}$, while for the Tsallis entropy (1.2), one has $H_q(f) = H_q(f_1) + H_q(f_2) + (1-q)H_q(f_1)H_q(f_2)$. The first property is known in physical literature as the extensivity property of Shannon and Rényi entropies, while the second is known as nonextensivity (with $q$ the parameter of nonextensivity).

The paper by Frank and Daffertshofer [10] presents a survey of results related to entropies in connection with nonlinear Fokker–Planck equations and normal or anomalous diffusion processes. In particular, the so-called Sharma and Mittal entropy $H_{q,s}=[1-(I_q)^{(s-1)/(q-1)}]/(s-1)$, with $q,s>0$, $q,s\neq 1$ and $I_q$ given by (1.4), represents a possible unification of the (nonextensive) Tsallis entropy (1.2) and (extensive) Rényi entropy (1.1). It satisfies $\lim_{s\to 1} H_{q,s}=H_q^*$, $\lim_{s,q\to 1} H_{q,s}=H_1$, $H_{q,q}=H_q$ and $\lim_{q\to 1} H_{q,s}=\{1-\exp[-(s-1)H_1]\}/(s-1)=H_s^G$, $s>0$, $s\neq 1$, where $H_s^G$ is known as Gaussian entropy. Notice that a consistent estimator of $H_{q,s}$ can be obtained from the estimator of $I_q$ presented in Section 3.

2.2. *Entropy maximizing distributions.* The $m$-dimensional random vector $X=([X]_1,\dots,[X]_m)^\top$ is said to have a multidimensional Student distribution $T(\nu,\Sigma,\mu)$ with mean $\mu\in\mathbb{R}^m$, scaling or correlation matrix $\Sigma$, covariance matrix $C=\nu\Sigma/(\nu-2)$ and $\nu$ degrees of freedom if its p.d.f. is

$$\begin{aligned} f_\nu(x) &= \frac{1}{(\nu\pi)^{m/2}} \\ &\quad\times \frac{\Gamma((m+\nu)/2)}{\Gamma(\nu/2)} \frac{1}{|\Sigma|^{1/2}[1+(x-\mu)^\top[\nu\Sigma]^{-1}(x-\mu)]^{(m+\nu)/2}}, \end{aligned} \tag{2.1}$$

$x\in\mathbb{R}^m$. The characteristic function of the distribution $T(\nu,\Sigma,\mu)$ is

$$\phi(\zeta)=\mathbb{E}\exp(i\langle\zeta,X\rangle)=\exp(i\langle\zeta,\mu\rangle)K_{\nu/2}(\sqrt{\nu\zeta^\top\Sigma\zeta})(\sqrt{\nu\zeta^\top\Sigma\zeta})^{\nu/2}\frac{2^{1-\nu/2}}{\Gamma(\nu/2)},$$

$\zeta\in\mathbb{R}^m$, where $K_{\nu/2}$ denotes the modified Bessel function of the second order. If $\nu=1$, then (2.1) is the $m$-variate Cauchy distribution. If $(\nu+m)/2$ is an integer, then (2.1) is the $m$-variate Pearson type VII distribution. If $Y$ is $\mathcal{N}(0,\Sigma)$ and if $\nu S^2$ is independent of $Y$ and $\mathcal{X}^2$-distributed with $\nu$ degrees of freedom, then $X=Y/S+\mu$ has the p.d.f. (2.1). The limiting form of (2.1) as $\nu\to\infty$ is the $m$-variate normal distribution $\mathcal{N}(\mu,\Sigma)$. The Rényi entropy (1.1) of (2.1) is

$$\begin{aligned} H_q^* &= \frac{1}{1-q}\log\frac{B(q(m+\nu)/2-m/2,m/2)}{B^q(\nu/2,m/2)} \\ &\quad+\frac{1}{2}\log[(\pi\nu)^m|\Sigma|]-\log\Gamma\left(\frac{m}{2}\right), \qquad q>\frac{m}{m+\nu}. \end{aligned}$$

It converges as $\nu\to\infty$ to the Rényi entropy

$$\begin{aligned} H_q^*(\mu,\Sigma) &= \log[(2\pi)^{m/2}|\Sigma|^{1/2}]-\frac{m}{2(1-q)}\log q \\ &= H_1(\mu,\Sigma)-\frac{m}{2}\left(1+\frac{\log q}{1-q}\right) \end{aligned} \tag{2.2}$$

of the multidimensional normal distribution $\mathcal{N}(\mu,\Sigma)$. When $q\to 1$, $H_q^*(\mu,\Sigma)$ tends to $H_1(\mu,\Sigma)=\log[(2\pi e)^{m/2}|\Sigma|^{1/2}]$, the Shannon entropy of $\mathcal{N}(\mu,\Sigma)$. For $m/(m+2)<q<1$, the $q$-entropy maximizing distribution under the constraint

$$\mathbb{E}(X-\mu)(X-\mu)^\top = C \tag{2.3}$$

is the Student distribution $T(\nu,(\nu-2)C/\nu,0)$ with $\nu=2/(1-q)-m>2$; see [43]. For $q>1$, we define $p=m+2/(q-1)$ and the $q$-entropy maximizing distribution under the constraint (2.3) has then finite support given by $\Omega_q=\{x\in\mathbb{R}^m:(x-\mu)^\top[(p+2)C]^{-1}(x-\mu)\le 1\}$. Its p.d.f. is

$$\begin{aligned} f_p(x) \\ = \begin{cases} \dfrac{\Gamma(p/2+1)}{|C|^{1/2}[\pi(p+2)]^{m/2}\Gamma((p-m)/2+1)} \\ \qquad \times [1-(x-\mu)^\top[(p+2)C]^{-1}(x-\mu)]^{1/(q-1)}, & \text{if } x\in\Omega_q \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.4}$$

The characteristic function of the p.d.f. (2.4) is given by

$$\phi(\zeta)=\exp(i\langle\zeta,\mu\rangle)2^{p/2}\Gamma\left(\frac{p}{2}+1\right)|\zeta^\top(p+2)C\zeta|^{-p/2}J_{p/2}(|\zeta^\top(p+2)C\zeta|),$$

$\zeta\in\mathbb{R}^m$, where $J_{\nu/2}$ denotes the Bessel function of the first kind.

Alternatively, $f_\nu$ for $q<1$ or $f_p$ for $q>1$ also maximizes the Shannon entropy (1.3) under a logarithmic constraint; see [20, 46]. Indeed, when $q<1$, $f_\nu(x)$ given by (2.1) with $\nu=2/(1-q)-m$ and $\Sigma=(\nu-1)C/\nu$ maximizes $H_1$ under the constraint

$$\int_{\mathbb{R}^m}\log(1+x^\top[(\nu-2)C]^{-1}x)f(x)\,dx=\Psi\left(\frac{\nu+m}{2}\right)-\Psi\left(\frac{\nu}{2}\right),$$

and when $q>1$, $f_p(x)$ given by (2.4) with $p=2/(q-1)+m$ maximizes $H_1$ under

$$\int_{\mathbb{R}^m}\log(1-x^\top[(p+2)C]^{-1}x)f(x)\,dx=\Psi\left(\frac{p}{2}\right)-\Psi\left(\frac{p+m}{2}\right),$$

where $\Psi(z)=\Gamma'(z)/\Gamma(z)$ is the digamma function.

2.3. *Information spectrum.* Considered as a function of $q$, $H_q^*$ (1.1) is known as the spectrum of Rényi information; see [36]. The value of $H_q^*$ for $q=2$ corresponds to the negative logarithm of the well-known efficacy parameter $\mathbb{E}f(X)$ that arises in asymptotic efficiency considerations. Consider now

$$\dot{H}_1=\lim_{q\to 1}\frac{dH_q^*}{dq}. \tag{2.5}$$

It satisfies

$$\begin{aligned}\dot{H}_1 &= \lim_{q\to 1} \frac{\log \int_{\mathbb{R}^m} f^q(x)\,dx}{(1-q)^2} + \frac{\int_{\mathbb{R}^m} f^q(x)\log f(x)\,dx}{(1-q)\int_{\mathbb{R}^m} f^q(x)\,dx}\\ &= -\frac{1}{2}\biggl\{\int_{\mathbb{R}^m} f(x)[\log f(x)]^2\,dx - \biggl[\int_{\mathbb{R}^m} f(x)\log f(x)\,dx\biggr]^2\biggr\}\\ &= -\frac{1}{2}\operatorname{var}[\log f(X)].\end{aligned}$$

The quantity $S(f) = -2\dot{H}_1 = \operatorname{var}[\log f(X)]$ gives a measure of the intrinsic shape of the density $f$; it is a location and scale invariant positive functional ($S(f) = S(g)$ when $f(x) = \sigma^{-1}g[(x-\mu)/\sigma]$). For the multivariate normal distribution $\mathcal{N}(\mu,\Sigma)$, $H_q^*$ is given by (2.2) and $S(f) = m/2$. For the one-dimensional Student distribution with $\nu$ degrees of freedom (for which $\mathbb{E}X^{\nu-1}$ exists, but not $\mathbb{E}X^{\nu}$), with density

$$f_\nu(x) = \frac{1}{(\nu\pi)^{1/2}} \frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)} \frac{1}{(1+x^2/\nu)^{(\nu+1)/2}},$$

we obtain

$$\begin{aligned}H_q^* &= \frac{1}{1-q}\log\frac{B(q(\nu+1)/2-1/2,1/2)}{B^q(\nu/2,1/2)} + \frac{1}{2}\log\nu, \qquad q > \frac{1}{\nu+1},\\ S(f_\nu) &= \begin{cases} \frac{\pi^2}{3} \simeq 3.2899, & \text{for } \nu = 1 \text{ (Cauchy distribution)},\\ 9 - \frac{3}{4}\pi^2 \simeq 1.5978, & \text{for } \nu = 2,\\ \frac{4}{3}\pi^2 - 12 \simeq 1.1595, & \text{for } \nu = 3,\\ \frac{775}{36} - \frac{25}{12}\pi^2 \simeq 0.9661, & \text{for } \nu = 4,\\ 3\pi^2 - \frac{115}{4} \simeq 0.8588, & \text{for } \nu = 5,\end{cases}\end{aligned} \tag{2.6}$$

and, more generally, $S(f_\nu) = (1/4)(\nu+1)^2\{\dot{\Psi}(\nu/2) - \dot{\Psi}[(\nu+1)/2]\}$, with $\dot{\Psi}(x)$ the trigamma function, $\dot{\Psi}(x) = d^2\log\Gamma(x)/dx^2$. The information provided by $S(f)$ on the shape of the distribution complements that given by other more classical characteristics like kurtosis. [Note that the kurtosis is not defined for $f_\nu$ when $\nu \le 4$; the one-dimensional Student distribution $f_6$ and the bi-exponential Laplace distribution $f_L$ have the same kurtosis but different values of $S(f)$ since $S(f_6) = 147931/3600 - (49/12)\pi^2 \simeq 0.7911$ and $S(f_L) = 1$.] For the multivariate Student distribution (2.1), we get $S(f_\nu) = (1/4)(\nu+m)^2\{\dot{\Psi}(\nu/2) - \dot{\Psi}[(\nu+m)/2]\}$. The $q$-entropy maximizing property of the Student distribution can be used to test that the observed samples are Student distributed, and the estimation of $S(f)$ then provides information about $\nu$. This finds important applications, for instance, in financial mathematics; see [18].

**3. Main results.** Let $\rho(x,y)$ denote the Euclidean distance between two points $x,y$ of $\mathbb{R}^m$ (see Section 5 for an extension to other metrics). For a given sample $X_1,\dots,X_N$, and a given $X_i$ in the sample, from the $N-1$ distances $\rho(X_i,X_j)$, $j=1,\dots,N$, $j\neq i$, we form the order statistics $\rho_{1,N-1}^{(i)}\le\rho_{2,N-1}^{(i)}\le\cdots\le\rho_{N-1,N-1}^{(i)}$. Therefore, $\rho_{1,N-1}^{(i)}$ is the nearest-neighbor distance from the observation $X_i$ to some other $X_j$ in the sample, $j\neq i$, and similarly, $\rho_{k,N-1}^{(i)}$ is the $k$th nearest-neighbor distance from $X_i$ to some other $X_j$.

3.1. *Rényi and Tsallis entropies.* We shall estimate $I_q$, $q\neq 1$, by

$$\hat{I}_{N,k,q}=\frac{1}{N}\sum_{i=1}^{N}(\zeta_{N,i,k})^{1-q}, \tag{3.1}$$

with

$$\zeta_{N,i,k}=(N-1)C_kV_m(\rho_{k,N-1}^{(i)})^m, \tag{3.2}$$

where $V_m=\pi^{m/2}/\Gamma(m/2+1)$ is the volume of the unit ball $\mathcal{B}(0,1)$ in $\mathbb{R}^m$ and

$$C_k=\left[\frac{\Gamma(k)}{\Gamma(k+1-q)}\right]^{1/(1-q)}.$$

Note that $I_1=1$ since $f$ is a p.d.f. and that $I_q$ is finite when $q<0$ only if $f$ is of bounded support. Indeed, $I_q=\int_{\{x:f(x)\ge1\}}f^q(x)\,dx+\int_{\{x:f(x)<1\}}f^q(x)\,dx>\int_{\{x:f(x)<1\}}f^q(x)\,dx>\mu_{\mathcal{L}}\{x:f(x)<1\}$, with $\mu_{\mathcal{L}}$ the Lebesgue measure. Also, when $f$ is bounded, $I_q$ tends to the (Lebesgue) measure of its support $\mu_{\mathcal{L}}\{x:f(x)>0\}$ when $q\to0^+$. Some other properties of $I_q$ are summarized in Lemma 1 of Section 6.

Remark 3.1. When $f$ is known, a Monte Carlo estimator of $I_q$ based on the sample $X_1,\dots,X_N$ is

$$\frac{1}{N}\sum_{i=1}^{N}f^{q-1}(X_i). \tag{3.3}$$

The nearest-neighbor estimator $\hat{I}_{N,k,q}$ given by (3.1) could thus also be considered as a plug-in estimator, $\hat{I}_{N,k,q}=(1/N)\sum_{i=1}^{N}[\hat{f}_{N,k}(X_i)]^{q-1}$, where $\hat{f}_{N,k}(x)=1/\{(N-1)C_kV_m[\rho_{k+1,N}(x)]^m\}$ with $\rho_{k+1,N}(x)$ the $(k+1)$th nearest-neighbor distance from $x$ to the sample. One may notice the resemblance between $\hat{f}_{N,k}(x)$ and the density function estimator $\tilde{f}_{N,k}(x)=k/\{NV_m[\rho_{k+1,N}(x)]^m\}$ suggested by Loftsgaarden and Quesenberry [26]; see also [7, 28].

We suppose that $X_1,\dots,X_N$, $N\geq 2$, are i.i.d. with a probability measure $\mu$ having a density $f$ with respect to the Lebesgue measure. [However, if $\mu$ has a finite number of singular components superimposed to the absolutely continuous component $f$, one can remove all zero distances from the $\rho_{k,N-1}^{(i)}$ in the computation of the estimate (3.1), which then enjoys the same properties as in Theorems 3.1 and 3.2, i.e., yields a consistent estimator of the Rényi and Tsallis entropies of the continuous component $f$.] The main results of the paper are as follows.

Theorem 3.1 (Asymptotic unbiasedness). *The estimator $\hat{I}_{N,k,q}$ given by (3.1) satisfies*

$$\mathbb{E}\hat{I}_{N,k,q}\to I_q, \qquad N\to\infty, \tag{3.4}$$

*for $q<1$, provided that $I_q$ given by (1.4) exists, and for any $q\in(1,k+1)$ if $f$ is bounded.*

Under the conditions of Theorem 3.1, $\mathbb{E}(1-\hat{I}_{N,k,q})/(q-1)\to H_q$ as $N\to\infty$, which provides an asymptotically unbiased estimate of the Tsallis entropy of $f$.

Theorem 3.2 (Consistency). *The estimator $\hat{I}_{N,k,q}$ given by (3.1) satisfies*

$$\hat{I}_{N,k,q}\xrightarrow{L_2} I_q, \qquad N\to\infty, \tag{3.5}$$

*(and thus, $\hat{I}_{N,k,q}\xrightarrow{\mathrm{p}} I_q$, $N\to\infty$) for $q<1$, provided that $I_{2q-1}$ exists, and for any $q\in(1,(k+1)/2)$ when $k\geq 2$ [resp. $q\in(1,3/2)$ when $k=1$] if $f$ is bounded.*

Corollary 3.1. *Under the conditions of Theorem 3.2,*

$$\hat{H}_{N,k,q}=(1-\hat{I}_{N,k,q})/(q-1)\xrightarrow{L_2} H_q \tag{3.6}$$

*and*

$$\hat{H}^*_{N,k,q}=\log(\hat{I}_{N,k,q})/(1-q)\xrightarrow{\mathrm{p}} H^*_q \tag{3.7}$$

*as $N\to\infty$, which provides consistent estimates of the Rényi and Tsallis entropies of $f$.*

We show the following in the proof of Theorem 3.2: when $q<1$ and $I_{2q-1}<\infty$, or $1<q<(k+2)/2$ and $f$ is bounded,

$$\mathbb{E}(\zeta_{N,i,k}^{1-q}-I_q)^2\to\Delta_{k,q}=I_{2q-1}\frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)}-I_q^2, \qquad N\to\infty.$$

Notice that $\lim_{k\to\infty}\Delta_{k,q} = I_{2q-1} - I_q^2 = \operatorname{var}[f^{q-1}(X)] = N\operatorname{var}[\frac{1}{N} \times \sum_{i=1}^N f^{q-1}(X_i)]$, that is, the limit of $\Delta_{k,q}$ for $k\to\infty$ equals $N$ times the variance of the Monte Carlo estimator (3.3) (which forms a lower bound on the variance of an estimator $I_q$ based on the sample $X_1,\ldots,X_N$).

Under the assumption that $f$ is three times continuously differentiable $\mu_{\mathcal{L}}$-almost everywhere, we can improve Lemma 2 of Section 6 into

$$\frac{1}{V_m R^m}\int_{\mathcal{B}(x,R)} f(z)\,dz = f(x) + \frac{R^2}{2(m+2)}\sum_{i=1}^m \frac{\partial^2 f(x)}{\partial x_i^2} + o(R^2), \qquad R\to 0,$$

which can be used to approximate $F_{N,x,k}(u) - F_{x,k}(u)$ in the proof of Theorem 3.1. We thereby obtain an approximation of the bias $\hat{B}_{N,k,q} = \mathbb{E}\hat{I}_{N,k,q} - I_q = \mathbb{E}\zeta_{N,1,k}^{1-q} - I_q$, which, after some calculations, can be written as

$$\hat{B}_{N,k,q} = \begin{cases} \dfrac{(q-1)(2-q)I_q}{2N} + \mathcal{O}(1/N^2), & \text{for } m=1,\\ \dfrac{q-1}{N}[(k+1-q)J_{q-2}/(8\pi) + (2-q)I_q/2] + \mathcal{O}(1/N^{3/2}), & \\ & \text{for } m=2,\\ \dfrac{q-1}{N^{2/m}}\dfrac{\Gamma(k+1+2/m-q)}{D_m\Gamma(k+1-q)}J_{q-1-2/m} + \mathcal{O}(1/N^{3/m}), & \\ & \text{for } m\ge 3,\end{cases}$$

where $J_\beta = \int_{\mathbb{R}^m} f^\beta(x)(\sum_{i=1}^m \partial^2 f(x)/\partial x_i^2)\,dx$ and $D_m = 2(m+2)V_m^{2/m}$. For instance, for $f$ the density of the normal $\mathcal{N}(0,\sigma^2 I_m)$, we get

$$J_\beta = -\frac{m}{\sigma^2}\frac{1}{(2\pi\sigma^2)^{m\beta/2}}\frac{\beta}{(\beta+1)^{1+m/2}},$$

which is defined for $\beta > -1$. From the expression of the MSE for $\hat{I}_{N,k,q}$ given in (6.8), we obtain

$$\mathbb{E}(\hat{I}_{N,k,q} - I_q)^2 = \frac{\Delta_{k,q}}{N} - 2I_q\hat{B}_{N,k,q}(1+o(1)) + [\mathbb{E}(\zeta_{N,1,k}^{1-q}\zeta_{N,2,k}^{1-q}) - I_q^2]. \tag{3.8}$$

Investigating the behavior of the last term requires an asymptotic approximation for $F_{N,x,y,k}(u,v) - F_{x,k}(u)F_{y,k}(v)$ (see the proof of Theorem 3.2), which is under current investigation. Preliminary results for $k=1$ show that the contribution of this term to the MSE for $\hat{I}_{N,k,q}$ cannot be ignored in general.

3.2. *Shannon entropy.* For the estimation of $H_1$ ($q=1$), we take the limit of $\hat{H}_{N,k,q}$ as $q\to 1$, which gives

$$\hat{H}_{N,k,1} = \frac{1}{N}\sum_{i=1}^N \log\xi_{N,i,k}, \tag{3.9}$$

with

$$\xi_{N,i,k} = (N-1)\exp[-\Psi(k)]V_m(\rho_{k,N-1}^{(i)})^m, \tag{3.10}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function [$\Psi(1) = -\gamma$ with $\gamma \simeq 0.5772$ the Euler constant and, for $k \geq 1$ integer, $\Psi(k) = -\gamma + A_{k-1}$ with $A_0 = 0$ and $A_j = \sum_{i=1}^{j} 1/i$]; see [22, 42] for applications of this estimator in physical sciences. We then have the following.

Corollary 3.2. *Suppose that $f$ is bounded and that $I_{q_1}$ exists for some $q_1 < 1$. Then $H_1$ exists and the estimator (3.9) satisfies $\hat{H}_{N,k,1} \overset{L_2}{\to} H_1$ as $N \to \infty$.*

Remark 3.2. One may notice that $\hat{H}^*_{N,k,q}$ given by (3.7) is a smooth function of $q$. Its derivative at $q = 1$ can be used as an estimate of $\dot{H}_1$ defined by (2.5). Straightforward calculations give

$$\begin{aligned}\lim_{q\to 1} \frac{d\hat{H}^*_{N,k,q}}{dq} &= \frac{\dot{\Psi}(k)}{2} - \frac{m^2}{2}\frac{1}{N}\sum_{i=1}^{N}\left[\log \rho_{k,N-1}^{(i)} - \frac{1}{N}\sum_{j=1}^{N}\log \rho_{k,N-1}^{(j)}\right]^2 \\ &= \frac{1}{2}\left[\dot{\Psi}(k) - \frac{1}{N}\sum_{i=1}^{N}(\log \xi_{N,i,k} - \hat{H}_{N,k,1})^2\right]\end{aligned}$$

and $S(f) = -2\dot{H}_1$ can be estimated by

$$\hat{S}_{N,k} = \frac{1}{N}\sum_{i=1}^{N}(\log \xi_{N,i,k} - \hat{H}_{N,k,1})^2 - \dot{\Psi}(k). \tag{3.11}$$

We obtain the following in the proof of Corollary 3.2:

$$\mathbb{E}(\log \xi_{N,i,k} - H_1)^2 \to \mathrm{var}[\log f(X)] + \dot{\Psi}(k), \qquad N \to \infty,$$

with $\dot{\Psi}(z) = d^2 \log \Gamma(z)/dz^2$ [and, for $k$ integer, $\dot{\Psi}(k) = \sum_{j=k}^{\infty} 1/j^2$]. Note that $\mathrm{var}[\log f(X)]$ forms a lower bound on the variance of a Monte Carlo estimation of $H_1$ based on $\log f(X_i)$, $i = 1, \ldots, N$, and that $\dot{\Psi}(k) \to 0$ as $k \to \infty$.

Similarly to Remark 3.1, the estimator $\hat{H}_{N,k,1}$ given by (3.9) could be considered as a plug-in estimator, $\hat{H}_{N,k,1} = -(1/N)\sum_{i=1}^{N} \log[\hat{f}'_{N,k}(X_i)]$ with $\hat{f}'_{N,k}(x) = \exp[\Psi(k)]/\{(N-1)V_m[\rho_{k+1,N}(x)]^m\}$. One may notice that selecting $k$ by likelihood cross-validation based on the density function estimator suggested by Loftsgaarden and Quesenberry [26], $\tilde{f}_{N,k}(x) = k/\{NV_m[\rho_{k+1,N}(x)]^m\}$, amounts to maximizing $-\hat{H}_{N,k,1} + \log k - \Psi(k)$, with $\log k - \Psi(k) = 1/(2k) + 1/(12k^2) + \mathcal{O}(1/k^4)$, $k \to \infty$. In our simulations this method always tended to select $k = 1$; replacing $\tilde{f}_{N,k}(x)$ by $\hat{f}'_{N,k}(x)$, or by $\hat{f}_{N,k}(x)$ of Remark 3.1, does not seem to yield a valid selection procedure for $k$ either.

Let $\tilde{H}_{N,k,1}$ be the plug-in estimator of $H_1$ based on $\tilde{f}_{N,k}$ defined by $\tilde{H}_{N,k,1} = -(1/N)\sum_{i=1}^{N} \log[\tilde{f}_{N,k}(X_i)]$. Then, under the conditions of Corollary 3.2, we obtain that $\lim_{N\to\infty} \mathbb{E}\tilde{H}_{N,k,1} = H_1 + \Psi(k) - \log k$ (since $\tilde{H}_{N,k,1} = \hat{H}_{N,k,1} + \Psi(k) - \log k + \log[N/(N-1)]$). Under the additional assumption on $f$ that it belongs to the class $\mathcal{F}$ of uniformly continuous p.d.f. satisfying $0 < c_1 \le f(x) \le c_2 < \infty$ for some constants $c_1, c_2$, we obtain the uniform and almost sure convergence of $\hat{H}_{N,k,1}$ to $H_1(f)$ over the class $\mathcal{F}$, provided that $k = k_N \to \infty$, $k_N/N \to 0$ and $k_N/\log N \to \infty$ as $N \to \infty$; see the results of Devroye and Wagner [7] on the strong uniform consistency of $\tilde{f}_{N,k}$. Notice that the choice of $k$ proposed by Hall, Park and Samworth [14] for nearest-neighbor classification does not satisfy these conditions.

3.3. *Relative entropy and divergences.* In some situations the statistical distance between distributions can be estimated through the computation of entropies, so that the method of $k$th nearest-neighbor distances presented above can be applied straightforwardly. For instance, the $q$-Jensen difference

$$J_q^\beta(f,g) = H_q^*[\beta f + (1-\beta)g] - [\beta H_q^*(f) + (1-\beta)H_q^*(g)], \qquad 0 \le \beta \le 1,$$

(see, e.g., [2]) can be estimated if we have three samples, respectively distributed according to $f$, $g$ and $\beta f + (1-\beta)g$. Suppose that we have one sample $S_i$ $(i = 1,\dots,s)$ of i.i.d. variables generated from $f$ and one sample $T_j$ $(j = 1,\dots,t)$ of i.i.d. variables generated from $g$ with $s$ and $t$ increasing at a constant rate as a function of $N = s + t$. Then, $H_q^*(f)$ and $H_q^*(g)$ can be estimated consistently from the two samples when $N \to \infty$; see Corollary 3.1. Also, as $N \to \infty$, the estimator $\hat{H}^*_{N,k,q}$ based on the sample $X_1,\dots,X_N$ with $X_i = S_i$ $(i = 1,\dots,s)$ and $X_i = T_{i-s}$ $(i = s+1,\dots,N)$ converges to $H_q^*[\beta f + (1-\beta)g]$, with $\beta = s/N$, and $J_q^\beta$ can therefore be estimated consistently from the two samples. This situation is encountered, for instance, in the image matching problem presented in [29], where entropy is estimated through the random graph approach of Redmond and Yukich [32]. As shown below, some other types of distances or divergences, that are not expressed directly through entropies, can also be estimated by the nearest-neighbor method.

Let $K(f,g)$ denote the Kullback–Leibler relative entropy,

$$K(f,g) = \int_{\mathbb{R}^m} f(x) \log \frac{f(x)}{g(x)}\, dx = \breve{H}_1 - H_1, \tag{3.12}$$

where $H_1$ is given by (1.3) and $\breve{H}_1 = -\int_{\mathbb{R}^m} f(x) \log g(x)\, dx$. Given $N$ independent observations $X_1,\dots,X_N$ distributed with the density $f$ and $M$ observations $Y_1,\dots,Y_M$ distributed with $g$, we wish to estimate $K(f,g)$.

The second term $H_1$ can be estimated by (3.9), with asymptotic properties given by Corollary 3.2. The first term $\breve{H}_1$ can be estimated in a similar manner, as follows: given $X_i$ in the sample, $i \in \{1,\dots,N\}$, consider $\breve{\rho}(X_i, Y_j)$, $j = 1,\dots,M$, and the order statistics $\breve{\rho}^{(i)}_{1,M} \le \breve{\rho}^{(i)}_{2,M} \le \cdots \le \breve{\rho}^{(i)}_{M,M}$, so that $\breve{\rho}^{(i)}_{k,M}$ is the $k$th nearest-neighbor distance from $X_i$ to some $Y_j$, $j \in \{1,\dots,M\}$. Then, one can prove, similarly to Corollary 3.2, that

$$\breve{H}_{N,M,k} = \frac{1}{N}\sum_{i=1}^{N} \log\{M \exp[-\Psi(k)] V_m (\breve{\rho}^{(i)}_{k,M})^m\} \tag{3.13}$$

is an asymptotically unbiased and consistent estimator of $\breve{H}_1$ (when now both $N$ and $M$ tend to infinity) when $g$ is bounded and

$$J_q = \int_{\mathbb{R}^m} f(x) g^{q-1}(x)\,dx \tag{3.14}$$

exists for some $q < 1$. The difference

$$\begin{aligned}\breve{H}_{N,M,k} - \hat{H}_{N,k,1} &= m \log\left[\prod_{i=1}^{N} \breve{\rho}^{(i)}_{k,M}\right]^{1/N} + \log M - \Psi(k) + \log V_m \\ &\quad - m \log\left[\prod_{i=1}^{N} \rho^{(i)}_{k,N}\right]^{1/N} - \log(N-1) + \Psi(k) - \log V_m \qquad (3.15)\\ &= m \log\left[\prod_{i=1}^{N} \frac{\breve{\rho}^{(i)}_{k,M}}{\rho^{(i)}_{k,N}}\right]^{1/N} + \log \frac{M}{N-1}\end{aligned}$$

thus gives an asymptotically unbiased and consistent estimator of $K(f, g)$. Obviously a similar technique can be used to estimate the (symmetric) Kullback–Leibler divergence $K(f, g) + K(g, f)$. Note, in particular, that when $f$ is unknown and only the sample $X_1,\dots,X_N$ is available while $g$ is known, then the term $\breve{H}_1$ in $K(f, g)$ can be estimated either by (3.13) with a sample $Y_1,\dots,Y_M$ generated from $g$, with $M$ taken arbitrarily large, or more simply by the Monte Carlo estimator

$$\breve{H}_{1,N}(g) = -\frac{1}{N}\sum_{i=1}^{N} \log g(X_i), \tag{3.16}$$

the term $H_1$ being still estimated by (3.9). This forms an alternative to the method by Broniatowski [6]. Compared to the method by Jiménez and Yukich [19] based on Voronoi tessellations (see also [27] for a Voronoi-based method for Shannon entropy estimation), it does not require any computation of multidimensional integrals. In some applications one wishes to optimize $K(f, g)$ with respect to $g$ that belongs to some class $G$ (possibly parametric), with $f$ fixed. Note that only the first term $\breve{H}_1$ of (3.12) then needs to be estimated. [Maximum likelihood estimation, with $g = g_\theta$ in a parametric class, is a most typical example: $\theta$ is then estimated by minimizing $\breve{H}_{1,N}(g_\theta)$; see (3.16).]

The Kullback–Leibler relative entropy can be used to construct a measure of mutual information (MI) between statistical distributions (see [22]) with applications in image [29, 44] and signal processing [23]. Let $a_i$ and $b_i$ denote the gray levels of pixel $i$ in two images $A$ and $B$ respectively, $i = 1, \ldots, N$. The image matching problem consists in finding an image $B$ in a data base that resembles a given reference image $A$. The MI method corresponds to maximizing $K(f, f_x f_y)$, with $f$ the joint density of the pairs $(a_i, b_i)$ and $f_x$ (resp. $f_y$) the density of gray levels in image $A$ (resp. $B$). We have $K(f, f_x f_y) = H_1(f_x) + H_1(f_y) - H_1(f)$, where each term can be estimated by (3.9) from one of the three samples $(a_i)$, $(b_i)$ or $(a_i, b_i)$ (but $A$ being fixed, only the last two terms need be estimated).

Another example of statistical distance between distributions is given by the following nonsymmetric Bregman distance

$$D_q(f,g) = \int_{\mathbb{R}^m} \left[ g^q(x) + \frac{1}{q-1} f^q(x) - \frac{q}{q-1} f(x) g^{q-1}(x) \right] dx, \quad q \neq 1, \tag{3.17}$$

or its symmetrized version

$$\begin{aligned} K_q(f,g) &= \frac{1}{q}[D_q(f,g) + D_q(g,f)] \\ &= \frac{1}{q-1} \int_{\mathbb{R}^m} [f(x) - g(x)][f^{q-1}(x) - g^{q-1}(x)]\, dx; \end{aligned}$$

see, for example, [2]. Given $N$ independent observations from $f$ and $M$ from $g$, the first and second terms in (3.17) can be estimated by using (3.1). In the last term, the integral $J_q$ given by (3.14) can be estimated by $\hat{I}_{N,M,k,q} = (1/N) \sum_{i=1}^N \{M C_k V_m (\breve{\rho}_{k,M}^{(i)})^m\}^{1-q}$. Similarly to Theorem 3.1, $\hat{I}_{N,M,k,q}$ is asymptotically unbiased, $N, M \to \infty$, for $q < 1$ if $J_q$ exists and for any $q \in (1, k+1)$ if $g$ is bounded. We also obtain a property similar to Theorem 3.2: $\hat{I}_{N,M,k,q}$ is a consistent estimator of $J_q$, $N, M \to \infty$, for $q < 1$ if $J_{2q-1}$ exists and for any $q \in (1, (k+2)/2)$ if $g$ is bounded. (Notice, however, the difference with Theorem 3.2: when $q > 1$ the cases $k = 1$ and $k \geq 2$ need not be distinguished for the estimation of $J_q$ and the upper bound on the admissible values for $q$ is slightly larger than in Theorem 3.2.)

## 4. Examples.

4.1. *Influence of $k$.* Figure 1 (left) presents $H_q^*$ as a function of $q$ (solid line) for the normal distribution $\mathcal{N}(0, I_3)$ in $\mathbb{R}^3$, together with estimates $\hat{H}_{N,k,q}^*$ for $k = 1, \ldots, 5$ obtained from a single sample of size $N = 1000$. Note that $\hat{H}_{N,k,q}^*$ is defined only for $q < k + 1$ and quickly deviates from the theoretical value $H_q^*$ when $q > (k+1)/2$ or $q < 1$ (the difficulties for $q$ small being due to $f$ having unbounded support). For comparison, we also compute a plug-in estimate of $H_q^*$ obtained through a (cross-validated) kernel density estimate of $f$. Define $\tilde{H}_{N,q}^* = \log(\tilde{I}_{N,q})/(1-q)$ and $\tilde{I}_{N,q} = (1/N)\sum_{i=1}^N \tilde{f}_{N,i}^{q-1}(X_i)$ with $\tilde{f}_{N,i}(x) = [(N-1)h^m(2\pi)^{m/2}]^{-1}\sum_{l=1,l\neq i}^N \exp\{-\|x - X_l\|^2/(2h^2)\}$, a $m$-variate cross-validated kernel estimator of $f$. No special care is taken for the choice of $h$ and we simply use the value that minimizes the asymptotic mean integrated squared error for the estimation of $f$, that is, $h = [4/(m+2)]^{1/(m+4)}N^{-1/(m+4)}$ with $m = 3$; see [34], page 152. The evolution of $\tilde{H}_{N,q}^*$ as a function of $q$ is plotted in dotted-line on Figure 1 (left): although the situation is favorable to kernel density estimation, $k$th nearest neighbors give a better estimation of $H_q^*$ for $q > 1$ and $k$ large enough. Figure 1 (right) shows $N$ times the empirical mean-squared error (MSE) $\mathbb{E}(\hat{H}_{N,k,q}^* - H_{N,q}^*)^2$ ($k = 1, 3, 5$) as a function of $q$ using 1 000 independent repetitions. The results for $N$ times the MSE $\mathbb{E}(\tilde{H}_{N,q}^* - H_{N,q}^*)^2$ for the

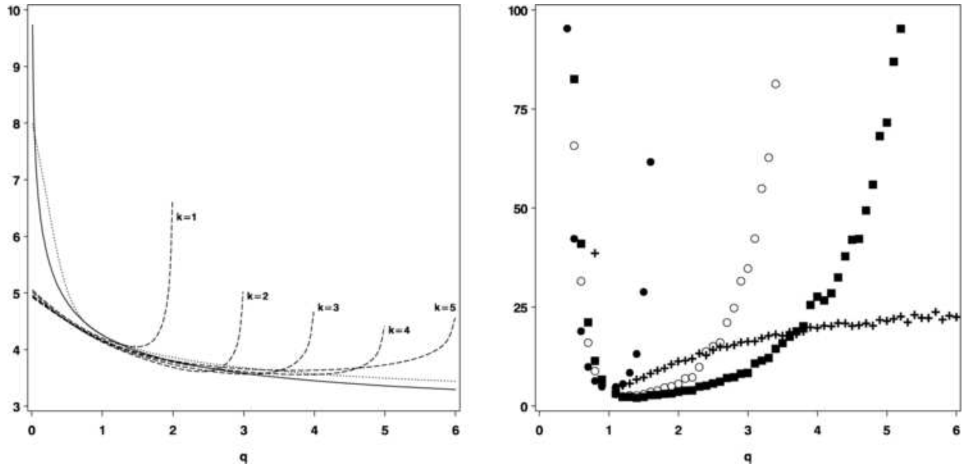


FIG. 1. *Behavior of estimators of entropy for samples from the normal distribution $\mathcal{N}(0, I_3)$ in $\mathbb{R}^3$ ($N = 1000$). [Left] $H_q^*$ (solid line), $\hat{H}_{N,k,q}^*$ (dashed lines) and $\tilde{H}_{N,q}^*$ obtained through a kernel estimation of $f$ (dotted line) as functions of $q$. [Right] $N = 1000$ times the empirical MSE for $\hat{H}_{N,k,q}$ [$k = 1$ (dots), $k = 3$ (circles), $k = 5$ (squares)] and for $\tilde{H}_{N,q}^*$ (plus) as a function of $q$ and computed over 1 000 independent samples.*

plug-in estimator are also shown. The figure indicates that the $k$th nearest neighbor estimator with $k$ satisfying $q<(k+1)/2$ is favorable in comparison to the plug-in estimator (for $q>1$ values of $k$ larger than 1 are preferable, whereas $k=1$ is preferable, for $q<1$).

Similar results hold for the Student distribution for $T(\nu,\Sigma,\mu)$ in $\mathbb{R}^3$ with 4 degrees of freedom, $\Sigma=I_3$ and $\mu=0$; see Figure 2. In selecting $k$ for $\hat{H}^*_{N,k,q}$, large values of $k$ are still generally preferable when $q>1$.

At this stage, the optimal selection of $k$ in $\hat{I}_{N,k,q}$ depending on $q$ and $N$ remains an open issue (see Sections 3.2 and 5). We repeated a series of intensive simulations to see how the MSE $\mathbb{E}(\hat{I}_{N,k,q}-I_q)^2$ evolves when $k$ varies, for different choices of $N$, $q$ and $m$. Figure 3 shows the influence of $N$ on the MSE for $\hat{I}_{N,k,q}$ for different values of $q$ using 10 000 independent repetitions, for $f$ the density of the standard normal $\mathcal{N}(0,1)$ and the normal $\mathcal{N}(0,I_3)$. For both $m=1$ and $m=3$ changes in $N$ appear to have a greater influence on $N$ times the MSE for $q=1.1$ in comparison to $q=4$. In particular, the figure indicates that for $m=3$ and $q=1.1$ the MSE decreases more slowly than $1/N$. Figure 4 shows the influence of $q$ on $N$ times the MSE for $\hat{I}_{N,k,q}$ as $k$ varies.

Although our simulations do not reveal a precise rule for choosing $k$, they indicate that this choice is not critical for practical applications: taking $k$ between 5 and 10 for $q\leq 2$ and increasing from 10 to 20 for $q$ from 2 to 4 gives reasonably good results for the cases we considered.

4.2. *Information spectrum, estimation of* $\mathrm{var}[\log f(X)]$. We use the method suggested in Remark 3.2 and estimate $S(f)=\mathrm{var}[\log f(X)]$ by $\hat{S}_{N,1}$ given

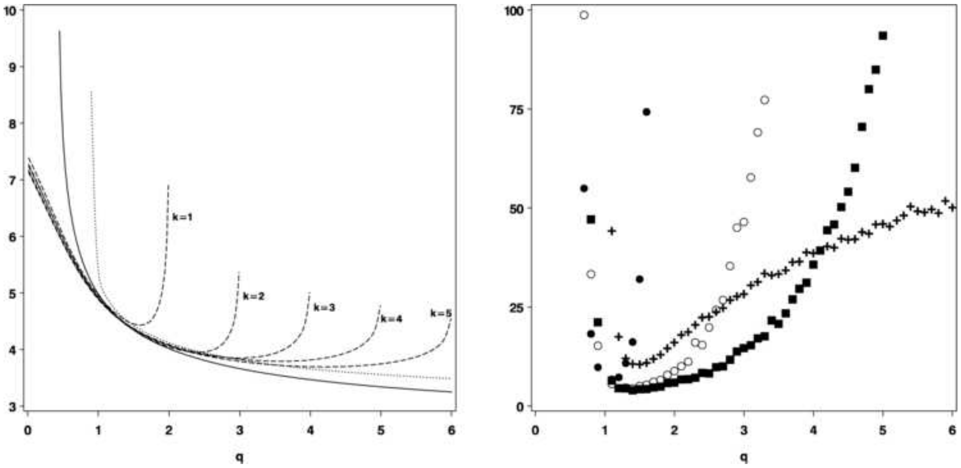
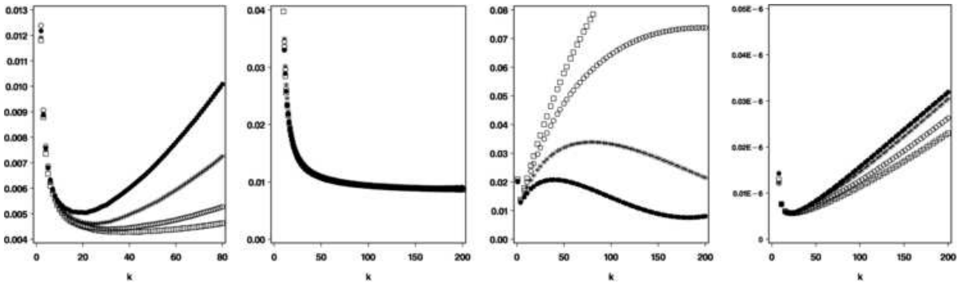


FIG. 2. *Same information as in Figure 1 but for the Student distribution $T(\nu,\Sigma,\mu)$ in $\mathbb{R}^3$ with 4 degrees of freedom ($\Sigma=I_3$, $\mu=0$, $N=1000$).*

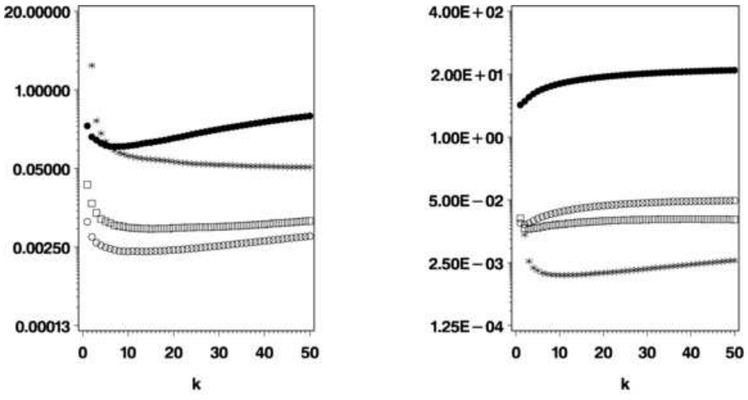


$q = 1.1$, $\mathcal{N}(0,1)$ $q = 4$, $\mathcal{N}(0,1)$ $q = 1.1$, $\mathcal{N}(0,I_3)$ $q = 4$, $\mathcal{N}(0,I_3)$

FIG. 3. *$N$ times the empirical MSE for $\hat{I}_{N,k,q}$ as a function of $k$ (10 000 independent repetitions), for $f$ the density of the standard normal $\mathcal{N}(0,1)$ and $\mathcal{N}(0,I_3)$ in $\mathbb{R}^3$ for varying $N$ $\{N = 1000$ (dots), 2 000 (stars), 5 000 (circles) and 10 000 (squares)$\}$ and $q = 1.1$ and $q = 4$.*

by (3.11) from a sample of 50 000 data generated with the Student distribution with 5 degrees of freedom. $S(f_\nu)$ is a decreasing function of $\nu$ and $S(f_4) \simeq 0.9661$, $S(f_5) \simeq 0.8588$, $S(f_6) \simeq 0.7911$; see Section 2.3. The empirical mean and standard deviation of $\hat{S}_{N,1}$ obtained from 10 000 independent repetitions are 0.8578 and 0.0269 respectively, indicating that $\nu$ can be correctly estimated in this way.

4.3. *Estimation of Kullback–Leibler divergence.* We use the same Student data as in 4.2 and estimate the Kullback–Leibler relative entropy $K(f, f_\nu)$ given by (3.12), using (3.16) for the estimation of $\check{H}_1$ and (3.9)



$\mathcal{N}(0,1)$ $\mathcal{N}(0,I_3)$

FIG. 4. *$N$ times the empirical MSE for $\hat{I}_{N,k,q}$ as a function of $k$ (10 000 independent repetitions), for $f$ the density of the standard normal $\mathcal{N}(0,1)$ and $\mathcal{N}(0,I_3)$ in $\mathbb{R}^3$ for varying $q$ $\{q = 0.75$ (dots), $q = 0.95$ (circles), $q = 1.1$ (squares) and $q = 2$ (stars)$\}$ and $N = 1000$.*

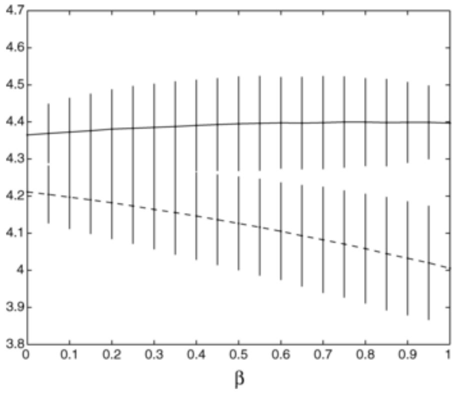


FIG. 5. *Empirical means of $\hat{H}^*_{N,3,0.75}$ (solid line) and $\hat{H}_{N,3,1}$ (dashed line) and two standard deviations (vertical bars) in a mixture of Student and normal distributions as functions of the mixture coefficient $\beta$ for $N = 500$ (1 000 independent repetitions).*

for the estimation of $H_1$, the entropy of $f$. The empirical means of the divergences estimated for $\nu = 1, \ldots, 8$ in 10 000 independent repetitions are 0.1657, 0.0440, 0.0119, 0.0021, 0.0000, 0.0012, 0.0038 and 0.0069 [the empirical standard deviations are rather large, approximately 0.0067 for each $\nu$, but the minimum is at $\nu = 5$ in all the 10 000 cases —notice that the dependence in $\nu$ is only through the term (3.16) where $f_\nu$ is substituted for $g$]. Again, $\nu$ is correctly estimated in this way.

4.4. *$q$-entropy maximizing distributions.* We generate $N = 500$ i.i.d. samples from the mixture of the three-dimensional Student distribution $T(\nu, (\nu - 2)/\nu I_3, 0)$ with $\nu = 5$ and the normal distribution $\mathcal{N}(0, I_3)$, with relative weights $\beta$ and $1 - \beta$. The covariance matrix of both distributions is the identity $I_3$, the Student distribution is $q$-entropy maximizing for $q = 1 - 2/(\nu + m) = 0.75$ (see Section 2.2) and the normal distribution maximizes Shannon entropy ($q = 1$). Figure 5 presents a plot of $\hat{H}^*_{N,k,q}$ for $q = 0.75$ and $\hat{H}_{N,k,1}$ as functions of the mixture coefficient $\beta$; both use $k = 3$ and are averaged over 1 000 repetitions, the vertical bars indicate two empirical standard deviations. [The values of $H^*_{0.75}$ estimated by plug-in using the kernel estimator $\tilde{f}_{N,i}(x)$ of Example 1 are totally out of the range for Student distributed variables due to the use of a nonadaptive bandwidth.]

**5. Related results and further developments.** The paper by Jiménez and Yukich [19] gives a method for estimating statistical distances between distributions with densities $f$ and $g$ based on Voronoi tessellations. Given an i.i.d. sample from $f$, it relies on the comparison between the Lebesgue measure (volume) and the measure for $g$ of the Voronoi cells (polyhedra) constructed from the sample. Voronoi tessellations are also used in [27] to estimate the Shannon entropy of $f$ based on an i.i.d. sample. The method requires the computation of the volumes of the Voronoi cells and no asymptotic result is given. Comparatively, the method based on nearest neighbors does not require any computation of (multidimensional) integrals. A possible motivation for using Voronoi tessellations could be the natural adaptation to the shape of the distribution. One may then notice that the metric used to compute nearest-neighbor distances can be adapted to the observed sample: for $X_1, \dots, X_N$, a sample having a nonspherical distribution, its empirical covariance matrix $\hat{\Sigma}_N$ can be used to define a new metric through $\|x\|^2_{\hat{\Sigma}_N} = x^\top \hat{\Sigma}_N^{-1} x$, the volume $V_m$ of the unit ball in this metric becoming $|\hat{\Sigma}_N|^{1/2}\pi^{m/2}/\Gamma(m/2+1)$.

$\sqrt{N}$-consistency of an estimator of $H_1$ based on nearest-neighbor distances ($k=1$) is proved by Tsybakov and van der Meulen [39] for $m=1$ and sufficiently regular densities $f$ with unbounded support using a truncation argument. On the other hand, $\sqrt{N}$-consistency of the estimator $\hat{I}_{N,k,q}$ is still an open issue (notice that the bias approximations of Section 3.1 indicate that it does not hold for large $m$). As for the case of spacing methods, where the spacing can be taken as an increasing function of the sample size $N$ (see, e.g., [12, 40, 41]) it might be of interest to let $k = k_N$ increase with $N$; see also [35] and Section 3.2. Properties of nearest-neighbor distances with $k_N \to \infty$ are considered, for instance, by Devroye and Wagner [7], Liero [24], Loftsgaarden and Quesenberry [26] and Moore and Yackel [28]. The derivation of an estimate of the asymptotic mean-squared error of the estimator could be used in a standard way to construct a rule for choosing $k$ as a function of $q$, $m$ and $N$ (see Sections 3.1 and 3.2). Numerical simulations indicate, however, that this choice is not as critical as that of the bandwidth in a kernel density estimator used for plug-in entropy estimation; see Section 4.

A central limit theorem for functions $h(\rho)$ of nearest-neighbor distances is obtained by Bickel and Breiman [4] for $k=1$ and by Penrose [30] for $k = k_N \to \infty$ as $N \to \infty$. However, their results do not apply to unbounded functions of $\rho$, such as $h(\rho) = \rho^{m(1-q)}$ [see (3.1)], or $h(\rho) = \log(\rho)$ [see (3.9)]. Conditions for the asymptotic normality of $\hat{I}_{N,k,q}$ are under current investigation.

## 6. Proofs.

The following lemma summarizes some properties of $I_q$.

Lemma 1.

(i) *If $f$ is bounded, then $I_q < \infty$ for any $q > 1$.*
(ii) *If $I_q < \infty$ for some $q < 1$, then $I_{q'} < \infty$ for any $q' \in (q, 1)$.*
(iii) *If $f$ is of finite support, $I_q < \infty$ for any $q \in [0, 1)$.*

PROOF.

(i) If $f(x) < \bar{f}$ and $q > 1$, $I_q = \int_{f \le 1} f^q + \int_{f>1} f^q \le \int_{f\le 1} f + \bar{f}^q \int_{f>1} f < \infty$.

(ii) If $q < q' < 1$, $I_{q'} = \int_{f\le 1} f^{q'} + \int_{f>1} f^{q'} \le \int_{f\le 1} f^q + \int_{f>1} f < \infty$ if $I_q < \infty$.

(iii) If $\mu_{\mathcal{S}} = \mu_{\mathcal{L}}\{x : f(x) > 0\} < \infty$ and $0 \le q < 1$, $I_q = \int_{f\le 1} f^q + \int_{f>1} f^q \le \mu_{\mathcal{S}} + \int_{f>1} f < \infty$. $\square$

The proofs of Theorems 3.1 and 3.2 use the following lemmas.

LEMMA 2 [Lebesgue (1910)]. *If $g \in L_1(\mathbb{R}^m)$, then for any sequence of open balls $\mathcal{B}(x, R_k)$ of radius tending to zero as $k \to \infty$ and for $\mu_{\mathcal{L}}$-almost any $x \in \mathbb{R}^m$,*

$$\lim_{k\to\infty} \frac{1}{V_m R_k^m} \int_{\mathcal{B}(x,R_k)} g(t)\,dt = g(x).$$

LEMMA 3. *For any $\beta > 0$,*

$$\int_0^\infty x^\beta F(dx) = \beta \int_0^\infty x^{\beta-1}[1 - F(x)]\,dx \tag{6.1}$$

*and*

$$\int_0^\infty x^{-\beta} F(dx) = \beta \int_0^\infty x^{-\beta-1} F(x)\,dx, \tag{6.2}$$

*in the sense that if one side converges so does the other.*

PROOF. See [9], volume 2, page 150, for (6.1). The proof is similar for (6.2). Define $\alpha = -\beta < 0$ and $I_{a,b} = \int_a^b x^\alpha F(dx)$ for some $a, b$, with $0 < a < b < \infty$. Integration by parts gives $I_{a,b} = [b^\alpha F(b) - a^\alpha F(a)] - \alpha \int_a^b x^{\alpha-1} F(x)\,dx$ and, since $\alpha < 0$, $\lim_{b\to\infty} I_{a,b} = I_{a,\infty} = -a^\alpha F(a) - \alpha \int_a^\infty x^{\alpha-1} F(x)\,dx < \infty$. Suppose that $\int_0^\infty x^{-\beta} F(dx) = J < \infty$. It implies $\lim_{a\to 0^+} I_{0,a} = 0$ and, since $I_{0,a} > a^\alpha F(a)$, $\lim_{a\to 0^+} a^\alpha F(a) = 0$. Therefore, $\lim_{a\to 0^+} -\alpha \int_a^\infty x^{\alpha-1} F(x)\,dx = J$.

Conversely, suppose that $\lim_{a\to 0^+} -\alpha \int_a^\infty x^{\alpha-1} F(x) dx = J < \infty$. Since $I_{a,\infty} < -\alpha \int_a^\infty x^{\alpha-1} F(x)\,dx$, $\lim_{a\to 0^+} I_{a,\infty} = J$. $\square$

6.1. *Proof of Theorem 3.1.* Since the $X_i$'s are i.i.d.,

$$\mathbb{E}\hat{I}_{N,k,q} = \mathbb{E}\zeta_{N,i,k}^{1-q} = \mathbb{E}[\mathbb{E}(\zeta_{N,i,k}^{1-q} | X_i = x)],$$

where the random variable $\zeta_{N,i,k}$ is defined by (3.2). Its distribution function conditional to $X_i = x$ is given by

$$F_{N,x,k}(u) = \Pr(\zeta_{N,i,k} < u | X_i = x) = \Pr[\rho_{k,N-1}^{(i)} < R_N(u) | X_i = x],$$

where

$$R_N(u) = \{u/[(N-1)V_m C_k]\}^{1/m}. \tag{6.3}$$

Let $\mathcal{B}(x,r)$ be the open ball of center $x$ and radius $r$. We have

$$\begin{aligned} F_{N,x,k}(u) &= \Pr\{k \text{ elements or more } \in \mathcal{B}[x, R_N(u)]\} \\ &= \sum_{j=k}^{N-1} \binom{N-1}{j} p_{N,u}^j (1-p_{N,u})^{N-1-j} \\ &= 1 - \sum_{j=0}^{k-1} \binom{N-1}{j} p_{N,u}^j (1-p_{N,u})^{N-1-j}, \end{aligned}$$

where $p_{N,u} = \int_{\mathcal{B}[x,R_N(u)]} f(t)\,dt$. From the Poisson approximation of binomial distribution, Lemma 2 gives

$$F_{N,x,k}(u) \to F_{x,k}(u) = 1 - \exp(-\lambda u) \sum_{j=0}^{k-1} \frac{(\lambda u)^j}{j!}$$

when $N \to \infty$ for $\mu$-almost any $x$, with $\lambda = f(x)/C_k$, that is, $F_{N,x,k}$ tends to the Erlang distribution $F_{x,k}$, with p.d.f. $f_{x,k}(u) = [\lambda^k u^{k-1} \exp(-\lambda u)]/\Gamma(k)$. Direct calculation gives

$$\int_0^\infty u^{1-q} f_{x,k}(u)\,du = \frac{\Gamma(k+1-q)}{\lambda^{1-q}\Gamma(k)} = f^{q-1}(x)$$

for any $q < k+1$.

Suppose first that $q < 1$ and consider the random variables $(U, X)$ with joint p.d.f. $f_{N,x,k}(u) f(x)$ on $\mathbb{R} \times \mathbb{R}^m$, where $f_{N,x,k}(u) = dF_{N,x,k}(u)/du$. The function $u \to u^{1-q}$ is bounded on every bounded interval and the generalized Helly–Bray Lemma (see [25], page 187) implies

$$\begin{aligned} \mathbb{E}\hat{I}_{N,k,q} &= \int_{\mathbb{R}^m} \int_0^\infty u^{1-q} f_{N,x,k}(u) f(x)\,du\,dx \\ &\to \int_{\mathbb{R}^m} f^q(x)\,dx = I_q, \qquad N \to \infty, \end{aligned}$$

which completes the proof.

Suppose now that $1 < q < k+1$. Note that from Lemma 1(i) $I_q < \infty$. Consider

$$J_N = \int_0^\infty u^{(1-q)(1+\delta)} F_{N,x,k}(du).$$

We show that $\sup_N J_N < \infty$ for some $\delta > 0$. From Theorem 2.5.1 of Bierens [5], page 34, it implies

$$z_{N,k}(x) = \int_0^\infty u^{1-q} F_{N,x,k}(du) \to z_k(x) = \int_0^\infty u^{1-q} F_{x,k}(du) = f^{q-1}(x), \quad N \to \infty \tag{6.4}$$

for $\mu$-almost any $x$ in $\mathbb{R}^m$.

Define $\beta = (1-q)(1+\delta)$, so that $\beta < 0$, and take $\delta < (k+1-q)/(q-1)$ so that $\beta + k > 0$. From (6.2),

$$\begin{aligned} J_N &= -\beta \int_0^\infty u^{\beta-1} F_{N,x,k}(u)\,du \\ &= -\beta \int_0^1 u^{\beta-1} F_{N,x,k}(u)\,du - \beta \int_1^\infty u^{\beta-1} F_{N,x,k}(u)\,du \\ &\le -\beta \int_0^1 u^{\beta-1} F_{N,x,k}(u)\,du - \beta \int_1^\infty u^{\beta-1}\,du \\ &= 1 - \beta \int_0^1 u^{\beta-1} F_{N,x,k}(u)\,du. \end{aligned} \tag{6.5}$$

Since $f(x)$ is bounded, say, by $\bar{f}$, we have $\forall x \in \mathbb{R}^m$, $\forall u \in \mathbb{R}$, $\forall N$, $p_{N,u} \le \bar{f} V_m [R_N(u)]^m = \bar{f} u / [(N-1)C_k]$. It implies

$$\begin{aligned} \frac{F_{N,x,k}(u)}{u^k} &\le \sum_{j=k}^{N-1} \binom{N-1}{j} \frac{\bar{f}^j u^{j-k}}{C_k^j (N-1)^j} \\ &\le \sum_{j=k}^{N-1} \frac{\bar{f}^j u^{j-k}}{C_k^j j!} = \frac{\bar{f}^k}{C_k^k k!} + \sum_{j=k+1}^{N-1} \frac{\bar{f}^j u^{j-k}}{C_k^j j!} \\ &\le \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \sum_{i=1}^{N-k-1} \frac{\bar{f}^i u^i}{C_k^i i!} \\ &\le \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \sum_{i=1}^{\infty} \frac{\bar{f}^i u^i}{C_k^i i!} = \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \left\{ \exp\left[ \frac{\bar{f} u}{C_k} \right] - 1 \right\}, \end{aligned}$$

and thus, for $u < 1$,

$$\frac{F_{N,x,k}(u)}{u^k} < U_k = \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \left\{ \exp\left[ \frac{\bar{f}}{C_k} \right] - 1 \right\}. \tag{6.6}$$

Therefore, from (6.5),

$$J_N \le 1 - \beta U_k \int_0^1 u^{k+\beta-1}\,du = 1 - \frac{\beta U_k}{k+\beta} < \infty, \tag{6.7}$$

which implies (6.4). Now we only need to prove that

$$\int_{\mathbb{R}^m} z_{N,k}(x) f(x)\,dx \to \int_{\mathbb{R}^m} z_k(x) f(x)\,dx = I_q, \qquad N \to \infty.$$

But this follows from Lebesgue's bounded convergence theorem, since $z_{N,k}(x)$ is bounded (take $\delta = 0$ in $J_N$).

6.2. *Proof of Theorem 3.2.* We shall use the same notations as in the proof of Theorem 3.1 and write $\hat{I}_{N,k,q} = (1/N)\sum_{i=1}^{N} \zeta_{N,i,k}^{1-q}$, so that

$$\begin{aligned}\mathbb{E}(\hat{I}_{N,k,q} - I_q)^2 &= \frac{\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2}{N} \\ &\quad + \frac{1}{N^2}\sum_{i \neq j} \mathbb{E}\{(\zeta_{N,i,k}^{1-q} - I_q)(\zeta_{N,j,k}^{1-q} - I_q)\}.\end{aligned} \tag{6.8}$$

We consider the cases $q < 1$ and $q > 1$ separately.

$q < 1$. Note that $2q - 1 < q < 1$ and Lemma 1(ii) gives $I_q < \infty$ when $I_{2q-1} < \infty$. Consider the first term on the right-hand side of (6.8). We have

$$\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2 = \mathbb{E}(\zeta_{N,i,k}^{1-q})^2 + I_q^2 - 2I_q\mathbb{E}\zeta_{N,i,k}^{1-q}, \tag{6.9}$$

where the last term tends to $-2I_q^2$ from Theorem 3.1. Consider the first term,

$$\mathbb{E}(\zeta_{N,i,k}^{1-q})^2 = \int_{\mathbb{R}^m}\int_0^\infty u^{2(1-q)} f_{N,x,k}(u) f(x)\,du\,dx.$$

Since the function $u \to u^{1-q}$ is bounded on every bounded interval, it tends to

$$\int_{\mathbb{R}^m}\int_0^\infty u^{2(1-q)} f_{x,k}(u) f(x)\,du\,dx = I_{2q-1}\frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)}$$

for any $q < (k+2)/2$ (generalized Helly–Bray lemma, Lóeve [25], page 187). Therefore, $\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2$ tends to a finite limit and the first term on the right-hand side of (6.8) tends to zero as $N \to \infty$.

Consider now the second term of (6.8). We show that

$$\begin{aligned}&\mathbb{E}\{(\zeta_{N,i,k}^{1-q} - I_q)(\zeta_{N,j,k}^{1-q} - I_q)\} \\ &\quad = \mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} + I_q^2 - 2I_q\mathbb{E}\zeta_{N,i,k}^{1-q} \to 0, \qquad N \to \infty.\end{aligned}$$

Since $\mathbb{E}\zeta_{N,i,k}^{1-q} \to I_q$ from Theorem 3.1, we only need to show that $\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to I_q^2$. Define

$$\begin{aligned}F_{N,x,y,k}(u,v) &= \Pr\{\zeta_{N,i,k} < u,\ \zeta_{N,j,k} < v | X_i = x, X_j = y\}, \\ &= \Pr\{\rho_{k,N-1}^{(i)} < R_N(u),\ \rho_{k,N-1}^{(j)} < R_N(v) | X_i = x, X_j = y\},\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \\ &\quad = \int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\int_0^\infty\int_0^\infty u^{1-q}v^{1-q} F_{N,x,y,k}(du,dv) f(x) f(y)\,dx\,dy.\end{aligned} \tag{6.10}$$

Let us assume that $x \neq y$. From the definition of $R_N(u)$ [see (6.3)] there exist $N_0 = N_0(x,y,u,v)$ such that $\mathcal{B}[x, R_N(u)] \cap \mathcal{B}[y, R_N(v)] = \varnothing$ for $N > N_0$ and thus,

$$F_{N,x,y,k}(u,v) = \sum_{j=k}^{N-2} \sum_{l=k}^{N-2-j} \binom{N-2}{j} \binom{N-2-j}{l} \\ \times p_{N,u}^j p_{N,v}^l (1 - p_{N,u} - p_{N,v})^{N-2-j-l}$$

with $p_{N,u} = \int_{\mathcal{B}[x,R_N(u)]} f(t)\,dt$, $p_{N,v} = \int_{\mathcal{B}[y,R_N(v)]} f(t)\,dt$. Hence, for $N > N_0$,

$$F_{N,x,y,k}(u,v) = F_{N-1,x,k}(u) + F_{N-1,y,k}(v) - 1 \\ + \sum_{j=0}^{k-1} \sum_{l=0}^{k-1} \binom{N-2}{j} \binom{N-2-j}{l} \\ \times p_{N,u}^j p_{N,v}^l (1 - p_{N,u} - p_{N,v})^{N-2-j-l}.$$

Similarly to the proof of Theorem 3.1, we then obtain

$$F_{N,x,y,k}(u,v) \to F_{x,y,k}(u,v) = F_{x,k}(u) F_{y,k}(v), \qquad N \to \infty, \tag{6.11}$$

for $\mu_{\mathcal{L}}$-almost any $x$ and $y$ with

$$\int_0^\infty \int_0^\infty u^{1-q} v^{1-q} F_{x,y,k}(du,dv) = f^{q-1}(x) f^{q-1}(y), \tag{6.12}$$

for any $q < k+1$. Since the function $u \to u^{1-q}$ is bounded on every bounded interval, (6.10) gives

$$\mathbb{E}\{\zeta_{N,i,k}^{1-q} \zeta_{N,j,k}^{1-q}\} \to \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f^q(x) f^q(y)\,dx\,dy = I_q^2, \qquad N \to \infty$$

(generalized Helly–Bray lemma, [25], page 187). This completes the proof that $\mathbb{E}(\hat{I}_{N,k,q} - I_q)^2 \to 0$. Therefore, $\hat{I}_{N,k,q} \xrightarrow{\text{p}} I_q$, when $N \to \infty$.

$q > 1$. Note that from Lemma 1(i) $I_q$ and $I_{2q-1}$ both exist. Consider the first term on the right-hand side of (6.8). We have again (6.9) where the last term tends to $-2I_q^2$ (the assumptions of the theorem imply $q < k+1$ so that Theorem 3.1 applies). Consider the first term of (6.9). Define

$$J'_N = \int_0^\infty u^{2(1-q)(1+\delta)} F_{N,x,k}(du),$$

we show that $\sup_N J'_N < \infty$ for some $\delta > 0$. From the assumptions of the theorem, $2q < k+2$. Let $\beta = 2(1-q)(1+\delta)$, so that $\beta < 0$ and take $\delta < (k+2-2q)/[2(q-1)]$ so that $\beta + k > 0$. Using Lemma 3 and developments similar to the proof of Theorem 3.1, we obtain

$$\begin{aligned} J'_N &= -\beta \int_0^\infty u^{\beta-1} F_{N,x,k}(du) \leq 1 - \beta \int_0^1 u^{\beta-1} F_{N,x,k}(du) \\ &\leq 1 - \beta U_k \int_0^1 u^{k+\beta-1}\, du = 1 - \frac{\beta U_k}{k+\beta} < \infty, \end{aligned}$$

where $U_k$ is given by (6.6). Theorem 2.5.1 of Bierens [5] then implies

$$\begin{aligned} \int_0^\infty u^{2(1-q)} F_{N,x,k}(du) &\to \int_0^\infty u^{2(1-q)} F_{x,k}(du) \\ &= \frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)} f^{2q-2}(x) \end{aligned}$$

for $\mu$-almost any $x$, $q < (k+2)/2$ and Lebesgue's bounded convergence theorem gives $\mathbb{E}(\zeta_{N,i,k}^{1-q})^2 \to I_{2q-1}\Gamma(k+2-2q)\Gamma(k)/\Gamma^2(k+1-q)$, $N \to \infty$. The first term of (6.8) thus tends to zero.

Consider now the second term. As in the case $q < 1$, we only need to show that $\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to I_q^2$. Define

$$J''_N = \int_0^\infty \int_0^\infty u^{(1-q)(1+\delta)} v^{(1-q)(1+\delta)} F_{N,x,y,k}(du, dv).$$

Using (6.11, 6.12), proving that $\sup_N J''_N < J(x,y) < \infty$ for some $\delta > 0$ will then establish that

$$\begin{aligned} &\int_0^\infty \int_0^\infty u^{1-q} v^{1-q} F_{N,x,y,k}(du, dv) \\ &\qquad \to f^{q-1}(x) f^{q-1}(y), \qquad N \to \infty, \end{aligned} \tag{6.13}$$

for $\mu$-almost $x$ and $y$; see Theorem 2.5.1 of Bierens [5]. Using (6.10), if

$$\int_{\mathbb{R}^m} \int_{\mathbb{R}^m} J(x,y) f(x) f(y)\, dx\, dy < \infty, \tag{6.14}$$

Lebesgue's dominated convergence theorem will then complete the proof.

Integration by parts, as in the proof of Lemma 3, gives

$$J''_N = \beta^2 \int_0^\infty \int_0^\infty u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\, du\, dv,$$

where $\beta = (1-q)(1+\delta) < 0$. We use different bounds for $F_{N,x,y,k}(u,v)$ on three different parts of the $(u,v)$ plane.

(i) Suppose that $\max[R_N(u), R_N(v)] \leq \|x-y\|$, which is equivalent to $(u,v) \in \mathcal{D}_1 = [0,\Lambda] \times [0,\Lambda]$ with $\Lambda = \Lambda(k,N,x,y) = (N-1)V_m C_k \|x-y\|^m$. This means that the balls $\mathcal{B}[x, R_N(u)]$ and $\mathcal{B}[y, R_N(v)]$ either do not intersect, or, when they do, their intersection contains neither $x$ nor $y$. In that case, we use

$$F_{N,x,y,k}(u,v) < \min[F_{N-1,x,k}(u), F_{N-1,y,k}(v)] < F_{N-1,x,k}^{1/2}(u) F_{N-1,y,k}^{1/2}(v)$$

and

$$\begin{aligned}
J_N''^{(1)} &= \beta^2 \int_{\mathcal{D}_1} u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\, du\, dv \\
&< \beta^2 \biggl[\int_0^\Lambda u^{\beta-1} F_{N-1,x,k}^{1/2}(u)\, du\biggr] \biggl[\int_0^\Lambda v^{\beta-1} F_{N-1,y,k}^{1/2}(v)\, dv\biggr] \\
&< \beta^2 \biggl[U_k^{1/2} \int_0^1 u^{\beta-1+k/2}\, du + \int_1^\infty u^{\beta-1}\, du\biggr]^2 \\
&= \beta^2 \biggl[U_k^{1/2} \frac{2}{2\beta+k} - \frac{1}{\beta}\biggr]^2 < \infty,
\end{aligned}$$

where we used the bound (6.6) for $F_{N-1,x,k}(u)$ when $u<1$, $F_{N-1,x,k}(u)<1$ for $u \geq 1$ and choose $\delta < (k+2-2q)/[2(q-1)]$ so that $2\beta+k>0$ [this choice of $\delta$ is legitimate since $q<(k+2)/2$].

(ii) Suppose, without any loss of generality, that $u<v$ and consider the domain defined by $R_N(u) \leq \|x-y\| < R_N(v)$, that is, $(u,v) \in \mathcal{D}_2 = [0,\Lambda] \times (\Lambda,\infty)$. The cases $k=1$ and $k \geq 2$ must be treated separately since $\mathcal{B}[y, R_N(v)]$ contains $x$.

When $k=1$, $F_{N,x,y,1}(u,v) = F_{N-1,x,1}(u)$ and we have

$$\begin{aligned}
J_N''^{(2)} &= \beta^2 \int_{\mathcal{D}_2} u^{\beta-1} v^{\beta-1} F_{N,x,y,1}(u,v)\, du\, dv \\
&< \beta^2 \biggl[\int_0^\Lambda u^{\beta-1} F_{N-1,x,1}(u)\, du\biggr] \biggl[\int_\Lambda^\infty v^{\beta-1}\, dv\biggr] \\
&< \beta^2 \biggl[U_1 \int_0^1 u^\beta du + \int_1^\infty u^{\beta-1}\, du\biggr] \biggl(-\frac{\Lambda^\beta}{\beta}\biggr) \qquad (6.15) \\
&= -\beta \biggl[\frac{U_1}{\beta+1} - \frac{1}{\beta}\biggr] \Lambda^\beta \\
&< J^{(2)}(x,y) = -\beta \biggl[\frac{U_1}{\beta+1} - \frac{1}{\beta}\biggr] V_m^\beta C_1^\beta \|x-y\|^{m\beta},
\end{aligned}$$

where we used (6.6) and take $\delta < (2-q)/(q-1)$ so that $\beta > -1$ (this choice of $\delta$ is legitimate since $q<2$).

Suppose now that $k \geq 2$. We have $F_{N,x,y,k}(u,v) < F_{N-1,x,k}^{1-\alpha}(u) F_{N-1,y,k-1}^\alpha(v)$, $\forall \alpha \in (0,1)$. Developments similar to those used for the derivation of (6.6) give for $v<1$

$$\begin{aligned}
&\frac{F_{N-1,y,k-1}(v)}{v^{k-1}} \\
&\quad < V_{k-1} = \frac{\bar{f}^{k-1}}{C_k^{k-1}(k-1)!} + \frac{\bar{f}^{k-1}}{C_k^{k-1}} \biggl\{\exp\biggl[\frac{\bar{f}}{C_k}\biggr] - 1\biggr\}. \qquad (6.16)
\end{aligned}$$

We obtain

$$\begin{aligned} J_N''^{(2)} &= \beta^2 \int_{\mathcal{D}_2} u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\,du\,dv \\ &< \beta^2 \left[\int_0^\Lambda u^{\beta-1} F_{N-1,x,k}^{1-\alpha}(u)\,du\right] \left[\int_\Lambda^\infty v^{\beta-1} F_{N-1,y,k-1}^{\alpha}(v)\,dv\right] \\ &< \beta^2 \left[U_k^{1-\alpha} \int_0^1 u^{\beta-1+(1-\alpha)k}\,du + \int_1^\infty u^{\beta-1}\,du\right] \\ &\quad \times \left[V_{k-1}^{\alpha} \int_0^1 v^{\beta-1+(k-1)\alpha}\,dv + \int_1^\infty v^{\beta-1}\,dv\right] \\ &= \beta^2 \left[\frac{U_k^{1-\alpha}}{k(1-\alpha)+\beta} - \frac{1}{\beta}\right] \left[\frac{V_{k-1}^{\alpha}}{(k-1)\alpha+\beta} - \frac{1}{\beta}\right] < \infty, \end{aligned}$$

where we used (6.6, 6.16) and require $\beta + k(1-\alpha) > 0$ and $\beta + (k-1)\alpha > 0$. For that we take $\alpha = \alpha_k = k/(2k-1)$. Indeed, from the assumptions of the theorem, $q < (k+1)/2 < (k^2+k-1)/(2k-1)$ so that we can choose $\delta < [(k^2+k-1) - q(2k-1)]/[(q-1)(2k-1)]$, which ensures that both $\beta + k(1-\alpha_k) > 0$ and $\beta + (k-1)\alpha_k > 0$.

(iii) Suppose finally that $\|x-y\| < \min[R_N(u), R_N(v)]$, that is, $(u,v) \in \mathcal{D}_3 = (\Lambda,\infty)\times(\Lambda,\infty)$. In that case, each of the balls $\mathcal{B}[x, R_N(u)]$ and $\mathcal{B}[y, R_N(v)]$ contains both $x$ and $y$. Again, the case $k=1$ and $k \geq 2$ must be distinguished.

When $k=1$, $F_{N,x,y,1}(u,v) = 1$ and

$$\begin{aligned} J_N''^{(3)} &= \beta^2 \int_{\mathcal{D}_3} u^{\beta-1} v^{\beta-1} F_{N,x,y,1}(u,v)\,du\,dv \\ &= \beta^2 \left[\int_\Lambda^\infty u^{\beta-1}\,du\right]^2 = \Lambda^{2\beta} \qquad (6.17) \\ &< J^{(3)}(x,y) = V_m^{2\beta} C_1^{2\beta} \|x-y\|^{2m\beta}. \end{aligned}$$

When $k \geq 2$, $F_{N,x,y,k}(u,v) < F_{N-1,x,k-1}^{1/2}(u) F_{N-1,y,k-1}^{1/2}(v)$ and

$$\begin{aligned} J_N''^{(3)} &= \beta^2 \int_{\mathcal{D}_3} u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\,du\,dv \\ &< \beta^2 \left[\int_\Lambda^\infty u^{\beta-1} F_{N-1,x,k-1}^{1/2}(u)\,du\right] \\ &\quad \times \left[\int_\Lambda^\infty v^{\beta-1} F_{N-1,y,k-1}^{1/2}(v)\,dv\right] \\ &< \beta^2 \left[V_{k-1}^{1/2} \frac{2}{2\beta+k-1} - \frac{1}{\beta}\right]^2 < \infty, \end{aligned}$$

where we used (6.16) and take $\delta < [(k+1)-2q]/[2(q-1)]$ so that $k-1+2\beta > 0$ [this choice of $\delta$ is legitimate since $q < (k+1)/2$].

Summarizing the three cases above, we obtain $J''_N = J''^{(1)}_N + 2J''^{(2)}_N + J''^{(3)}_N$ with different bounds for $J''^{(2)}_N$ and $J''^{(3)}_N$ depending on whether $k = 1$ or $k \geq 2$. This proves (6.13).

When $k \geq 2$, the bound on $J''_N$ does not depend on $x, y$ and Lebesgue's bounded convergence theorem implies $\mathbb{E}\{\zeta^{1-q}_{N,i,k}\zeta^{1-q}_{N,j,k}\} \to I^2_q$, which completes the proof of the theorem; see (6.14).

When $k = 1$, the condition (6.14) is satisfied if $2\beta > -1$ [see (6.15), (6.17)], which is ensured by the choice $\delta < (3-2q)/[2(q-1)]$ (legitimate since $q < 3/2$). Indeed, we can write

$$\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \|x-y\|^\gamma f(x)f(y)\,dx\,dy = \int_{\mathbb{R}^m} \|x\|^\gamma g(x)\,dx,$$

where $g(x) = \int_{\mathbb{R}^m} f(x+y)f(y)\,dy$, and thus (since $\gamma < 0$),

$$\begin{aligned}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \|x-y\|^\gamma f(x)f(y)\,dx\,dy &\leq \bar{f}^2 \int_{\|x\|<1} \|x\|^\gamma dx + I_2 \\ &= \bar{f}^2 \frac{mV_m}{\gamma+m} + I_2,\end{aligned}$$

when $\gamma > -m$. When $\delta < (3-2q)/[2(q-1)]$, Lebesgue's dominated convergence theorem thus implies $\mathbb{E}\{\zeta^{1-q}_{N,i,k}\zeta^{1-q}_{N,j,k}\} \to I^2_q$ , which completes the proof of the theorem.

6.3. *Proof of Corollary 3.2.* The existence of $H_1$ directly follows from that of $I_{q_1}$ for $q_1 < 1$ and the boundedness of $f$.

*Asymptotic unbiasedness.* We have

$$\mathbb{E}\hat{H}_{N,k,1} = \mathbb{E}\log\xi_{N,i,k} = \mathbb{E}[\mathbb{E}(\log\xi_{N,i,k}|X_i = x)],$$

where the only difference between the random variables $\zeta_{N,i,k}$ (3.10) and $\xi_{N,i,k}$ (3.2) is the substitution of $\exp[-\Psi(k)]$ for $C_k$. Similarly to the proof of Theorem 3.1, we define $F_{N,x,k}(u) = \Pr(\xi_{N,i,k} < u|X_i = x) = \Pr[\rho^{(i)}_{k,N-1} < R_N(u)|X_i = x]$ with now $R_N(u) = (u/\{(N-1)V_m\exp[-\Psi(k)]\})^{1/m}$. Following the same steps as in the proof of Theorem 3.1, we then obtain

$$F_{N,x,k}(u) \to F_{x,k}(u) = 1 - \exp(-\lambda u)\sum_{j=0}^{k-1}\frac{(\lambda u)^j}{j!}, \qquad N \to \infty,$$

for $\mu_{\mathcal{L}}$-almost any $x$, with $\lambda = f(x)\exp[\Psi(k)]$.

Direct calculation gives $\int_0^\infty \log(u) F_{x,k}(du) = -\log f(x)$. We shall use again Theorem 2.5.1 of Bierens [5], page 34, and show that

$$J_N = \int_0^\infty |\log(u)|^{1+\delta} F_{N,x,k}(du) < \infty, \tag{6.18}$$

for some $\delta > 0$, which implies

$$\int_0^\infty \log(u) F_{N,x,k}(du) \to \int_0^\infty \log(u) F_{x,k}(du) = -\log f(x), \qquad N \to \infty,$$

for $\mu_{\mathcal{L}}$-almost any $x$. The convergence

$$\int_{\mathbb{R}^m} \int_0^\infty \log(u) F_{N,x,k}(du) f(x)\, dx \to H_1, \qquad N \to \infty,$$

then follows from Lebesgue's bounded convergence theorem.

In order to prove (6.18), we write

$$J_N = \int_0^1 |\log(u)|^{1+\delta} F_{N,x,k}(du) + \int_1^\infty |\log(u)|^{1+\delta} F_{N,x,k}(du). \tag{6.19}$$

Since $f$ is bounded, we can take $q_2 > 1$ (and smaller than $k+1$) such that $\int_0^\infty u^{1-q_2} F_{N,x,k}(du) < \infty$; see (6.7). Since $|\log(u)|^{1+\delta}/u^{1-q_2} \to 0$ when $u \to 0$, it implies that the first integral on the right-hand side of (6.19) is finite. Similarly, since, by assumption, $I_{q_1}$ exists for some $q_1 < 1$, $\int_0^\infty u^{1-q_1} F_{N,x,k}(du) < \infty$ and $|\log(u)|^{1+\delta}/u^{1-q_1} \to 0$, $u \to \infty$, implies that the second integral on the right-hand side of (6.19) is finite, which completes the proof that $\mathbb{E}\hat{H}_{N,k,1} \to H_1$ as $N \to \infty$.

*$L_2$ consistency.* Similarly to the proof of asymptotic unbiasedness, we only need to replace $\zeta_{N,i,k}$ (3.10) by $\xi_{N,i,k}$ (3.2) and $C_k$ by $\exp[-\Psi(k)]$ in the proof of Theorem 3.2. When we now compute

$$\begin{aligned} \mathbb{E}(\hat{H}_{N,k,1} - H_1)^2 &= \frac{\mathbb{E}(\log \xi_{N,i,k} - H_1)^2}{N} \\ &\quad + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E}\{(\log \xi_{N,i,k} - H_1)(\log \xi_{N,j,k} - H_1)\}, \end{aligned} \tag{6.20}$$

in the first term, $\mathbb{E}(\log \xi_{N,i,k} - H_1)^2$ tends to

$$\int_{\mathbb{R}^m} \log^2 f(x) f(x)\, dx - H_1^2 + \dot{\Psi}(k) = \operatorname{var}[\log f(X)] + \dot{\Psi}(k),$$

where $\dot{\Psi}(z)$ is the trigamma function, $\dot{\Psi}(z) = d^2 \log \Gamma(z)/dz^2$, and for the second term the developments are similar to those in Theorem 3.2. For instance, equation (6.13) now becomes $\int_0^\infty \int_0^\infty \log u \log v F_{N,x,y,k}(du, dv) \to \log f(x) \log f(y)$, $N \to \infty$, for $\mu$-almost $x$ and $y$. We can then show that $\mathbb{E}\{\log \xi_{N,i,k} \log \xi_{N,j,k}\} \to H_1^2$, so that $\mathbb{E}(\hat{H}_{N,k,1} - H_1)^2 \to 0$, $N \to \infty$.

**Acknowledgments.** The authors wish to thank Anatoly A. Zhigljavsky from the Cardiff School of Mathematics for helpful discussions. Comments from the associate editor and the referees helped in improving the presentation of our results.

## REFERENCES

[1] Alemany, P. and Zanette, S. (1992). Fractal random walks from a variational formalism for Tsallis entropies. *Phys. Rev. E* **49** 956–958.

[2] Basseville, M. (1996). Information entropies, divergences et moyennes. Research Report IRISA nb. 1020.

[3] Beirlant, J., Dudewicz, E. J., Györfi, L. and van der Meulen, E. C. (1997). Nonparametric entropy estimation: An overview. *Internat. J. Math. Statist. Sci.* **6** 17–39. MR1471870

[4] Bickel, P. J. and Breiman, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214. MR0682809

[5] Bierens, H. J. (1994). *Topics in Advanced Econometrics*. Cambridge Univ. Press. MR1291390

[6] Broniatowski, M. (2003). Estimation of the Kullback–Leibler divergence. *Math. Methods Statist.* **12** 391–409. MR2054155

[7] Devroye, L. P. and Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5** 536–540. MR0436442

[8] Evans, D., Jones, A. J. and Schmidt, W. M. (2002). Asymptotic moments of near-neighbour distance distributions. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **458** 2839–2849. MR1987515

[9] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*. **II**. Wiley, New York. MR0210154

[10] Frank, T. and Daffertshofer, A. (2000). Exact time-dependent solutions of the Rényi Fokker–Planck equation and the Fokker–Planck equations related to the entropies proposed by Sharma and Mittal. *Phys. A* **285** 351–366.

[11] Goria, M. N., Leonenko, N. N., Mergel, V. V. and Novi Inverardi, P. L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametr. Statist.* **17** 277–297. MR2129834

[12] Hall, P. (1986). On powerful distributional tests based on sample spacings. *J. Multivariate Statist.* **19** 201–225. MR0853053

[13] Hall, P. and Morton, S. (1993). On the estimation of entropy. *Ann. Inst. Statist. Math.* **45** 69–88. MR1220291

[14] Hall, P., Park, B. and Samworth, R. (2004). Choice of neighbour order in nearest-neighbour classification. Manuscript.

[15] Havrda, J. and Charvát, F. (1967). Quantification method of classification processes. Concept of structural $\alpha$-entropy. *Kybernetika* (*Prague*) **3** 30–35. MR0209067

[16] Hero, III, A. O., Ma, B., Michel, O. and Gorman, J. (2002). Applications of entropic spanning graphs. *IEEE Signal. Proc. Magazine* **19** 85–95. (Special Issue on Mathematics in Imaging.)

[17] Hero, III, A. O. and Michel, O. J. J. (1999). Asymptotic theory of greedy approximations to minimal $k$-point random graphs. *IEEE Trans. Inform. Theory* **45** 1921–1938. MR1720641

[18] Heyde, C. C. and Leonenko, N. N. (2005). Student processes. *Adv. in Appl. Probab.* **37** 342–365. MR2144557

[19] Jiménez, R. and Yukich, J. E. (2002). Asymptotics for statistical distances based on Voronoi tessellations. *J. Theoret. Probab.* **15** 503–541. MR1898817

[20] Kapur, J. N. (1989). *Maximum-Entropy Models in Science and Engineering*. Wiley, New York. MR1079544

[21] Kozachenko, L. and Leonenko, N. (1987). On statistical estimation of entropy of a random vector. *Problems Inform. Transmission* **23** 95–101. [Translated from *Prolemy Predachi Informatsii* **23** (1987) 9–16 (in Russian).] MR0908626

[22] Kraskov, A., Stögbauer, H. and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* **69** 1–16. MR2096503

[23] Learned-Miller, E. and Fisher, J. (2003). ICA using spacings estimates of entropy. *J. Machine Learning Research* **4** 1271–1295. MR2103630

[24] Liero, H. (1993). A note on the asymptotic behaviour of the distance of the $k$th nearest neighbour. *Statistics* **24** 235–243. MR1240938

[25] Loève, M. (1977). *Probability Theory I*, 4th ed. Springer, New York. MR0651017

[26] Loftsgaarden, D. O. and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36** 1049–1051. MR0176567

[27] Miller, E. (2003). A new class of entropy estimators for multidimensional densities. In *Proc. ICASSP'2003*.

[28] Moore, D. S. and Yackel, J. W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.* **5** 143–154. MR0426275

[29] Neemuchwala, H., Hero, A. and Carson, P. (2005). Image matching using alpha-entropy measures and entropic graphs. *Signal Processing* **85** 277–296.

[30] Penrose, M. D. (2000). Central limit theorems for $k$-nearest neighbour distances. *Stochastic Process. Appl.* **85** 295–320. MR1731028

[31] Pronzato, L., Thierry, É. and Wolsztynski, É. (2004). Minimum entropy estimation in semi-parametric models: A candidate for adaptive estimation? In *mODa 7—Advances in Model-Oriented Design and Analysis. Contrib. Statist.* 125–132. Physica, Heidelberg. MR2089333

[32] Redmond, C. and Yukich, J. E. (1996). Asymptotics for Euclidean functionals with power-weighted edges. *Stochastic Process. Appl.* **61** 289–304. MR1386178

[33] Rényi, A. (1961). On measures of entropy and information. *Proc. 4th Berkeley Sympos. Math. Statist. Probab.* **I** 547–561. Univ. California Press, Berkeley. MR0132570

[34] Scott, D. W. (1992). *Multivariate Density Estimation*. Wiley, New York. MR1191168

[35] Song, K. (2000). Limit theorems for nonparametric sample entropy estimators. *Statist. Probab. Lett.* **49** 9–18. MR1789659

[36] Song, K. (2001). Rényi information, loglikelihood and an intrinsic distribution measure. *J. Statist. Plann. Inference* **93** 51–69. MR1822388

[37] Tsallis, C. (1988). Possible generalization of Boltzmann–Gibbs statistics. *J. Statist. Phys.* **52** 479–487. MR0968597

[38] Tsallis, C. and Bukman, D. (1996). Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis. *Phys. Rev. E* **54** 2197–2200.

[39] Tsybakov, A. B. and van der Meulen, E. C. (1996). Root-$n$ consistent estimators of entropy for densities with unbounded support. *Scand. J. Statist.* **23** 75–83. MR1380483

[40] van Es, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.* **19** 61–72. MR1172967

[41] Vasicek, O. (1976). A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* **38** 54–59. MR0420958

[42] Victor, J. (2002). Binless strategies for estimation of information from neural data. *Phys. Rev. E* **66** 1–15.

[43] Vignat, C., Hero, III, A. O. and Costa, J. A. (2004). About closedness by convolution of the Tsallis maximizers. *Phys. A* **340** 147–152. MR2088335

[44] Viola, P. and Wells, W. (1995). Alignment by maximization of mutual information. In *5th IEEE Internat. Conference on Computer Vision* 16–23. Cambridge, MA.

[45] Wolsztynski, É., Thierry, É. and Pronzato, L. (2005). Minimum entropy estimation in semi-parametric models. *Signal Processing* **85** 937–949.

[46] Zografos, K. (1999). On maximum entropy characterization of Pearson's type II and VII multivariate distributions. *J. Multivariate Anal.* **71** 67–75. MR1721960

N. Leonenko
V. Savani
Cardiff School of Mathematics
Cardiff University
Senghennydd Road
Cardiff CF24 4AG
United Kingdom
E-mail: leonenkon@cardiff.ac.uk
savaniv@cardiff.ac.uk

L. Pronzato
Laboratoire I3S
Les Algorithmes
CNRS/Université de Nice–Sophia Antipolis
2000 route des Lucioles
BP 121
06903 Sophia-Antipolis Cedex
France
E-mail: pronzato@i3s.unice.fr